Factor exposures and diversification: are sustainably screened portfolios any different?

Arnaud Gougler, Sebastian Utz

Angaben zur Veröffentlichung / Publication details:

Gougler, Arnaud, and Sebastian Utz. 2020. "Factor exposures and diversification: are sustainably screened portfolios any different?" *Financial Markets and Portfolio Management* 34 (3): 221–49. https://doi.org/10.1007/s11408-020-00354-4.

Factor exposures and diversification: Are sustainably screened portfolios any different?

Arnaud Gougler[1] · **Sebastian Utz**[2]

Abstract
We analyze the performance, risk, and diversification characteristics of global screened and best-in-class equity portfolios constructed according to Inrate's sustainability ratings. The financial performance of sustainably high-rated portfolios is similar to the risk-adjusted market performance in terms of abnormal returns of a five-factor market model. In contrast, low-rated portfolios exhibit negative abnormal returns. Firms with high sustainability ratings show lower idiosyncratic risk and higher exposure toward the high-minus-low and the conservative-minus-aggressive factor.

Keywords Sustainable portfolios · Portfolio diversification · ESG scores · Screening approaches · Idiosyncratic risk

JEL Classification G11 · Q56

1 Introduction

Sustainable investments refer to investments that incorporate environmental, social, and governance (ESG) criteria into investment decisions (Laurence 2013). Over the past decades, such investments have become a major trend worldwide and continue to grow at a steady rate (Eurosif 2018; Renneboog et al. 2008). This evolution has made sustainable investments a key topic in financial research (see Auer and Schuhmacher 2016; Derwall et al. 2011; Renneboog et al. 2008; Utz and Wimmer 2014; Schröder 2004; Walker et al. 2014), and various aspects of these types of investments have been scrutinized such as their performance, their risk characteristics, or the real sustainabil-

✉ Sebastian Utz
sebastian.utz@unisg.ch

Arnaud Gougler
arnaud.gougler@unifr.ch

1 University of Fribourg, 1700 Fribourg, Switzerland

2 School of Finance, University St. Gallen, Bodanstrasse 6, 9000 St. Gallen, Switzerland

ity impact of mutual funds labeled as sustainable. In this study, we build on existing financial markets and portfolio management literature on different kinds of sustainable screenings (Areal et al. 2013) and investigate whether sustainably screened portfolios exhibit diversification characteristics that are different from those of unscreened portfolios. Furthermore, we investigate whether portfolios with different screens based on sustainability ratings display different investment styles such as different book-to-market ratios (see Derwall et al. 2005; Galema et al. 2008). We employ ESG data of the Swiss sustainability rating agency Inrate to apply different levels of screening intensity. We show that portfolios comprising sustainable firms exhibit similar risk-adjusted returns and risk measures, yet diverging investment styles and increased diversification measures as market portfolios.

Both in academia and practice, sustainable investments are considered cautiously due to a variety of obscurities concerning implementation, risk, performance, and impact measurement. Indeed, a skeptic strand of the literature argues that, according to the neoclassical economics view, firms' role is to maximize profits. The pursuit of objectives different than profit maximization, such as sustainability, allocates resources to deviating goals and ultimately leads to a suboptimal bottom line (Barnett and Salomon 2006; Carroll and Shabana 2010; Renneboog et al. 2008). In contrast, proponents of the benefits of sustainable investments put forward financially driven motivation that includes risk management and long-term return (see, e.g., Bouslah et al. 2013; Harjoto and Laksmana 2018). According to this view, firms with high sustainable scores display enhanced efficiency, achieve differentiation on the market, avoid regulatory pressure and costs, avoid or minimize litigation risks from stakeholders, experience improved relationships with stakeholders, and have sound management practices (Barnett and Salomon 2006; Carroll and Shabana 2010; Verheyden et al. 2016; de Carvalho Ferreira et al. 2016; Renneboog et al. 2008). Furthermore, the proactive approach toward risks leads to the avoidance of high scandal or crisis costs (Kim et al. 2014; Utz 2018). These characteristics might thus translate into lower risk that offsets possibly lower returns in a risk-adjusted performance perspective.

In our study, we use Inrate's sustainability ratings to determine the level of sustainability of firms. We integrate sustainability into investment portfolios following the two most commonly applied strategies in sustainable investment (e.g., see, Renneboog et al. 2008; Scholtens 2006): negative screens (i.e., excluding controversial firms and industries from the investment universe) and best-in-class (i.e., concentrating on a certain proportion of best-performing firms in each sector concerning sustainable criteria). In the first step, we calculate different performance, risk, and diversification measures for screened and best-in-class portfolios. Due to data availability, the analysis period spans between October 2013 and May 2017. While we find abnormal returns similar to zero for both types of portfolios in general, moderate levels of negative screens generate abnormal returns that substantially differ from those of the unrestricted universe. Both risk-adjusted performance and idiosyncratic risk increase for screened portfolios with increasing selection intensity. In particular, portfolios investing in firms of the highest five sustainability rating levels perform best over the period in terms of diversification and risk-adjusted performance and show metrics that are significantly higher than the unrestricted universe. Except for the idiosyncratic risk, the performance metrics of best-in-class portfolios remain similar to unscreened portfolios even for the most sustainable portfolios (best 10%).

We continue with an analysis of subsamples ("A+" to "D−") of Inrate's classification to identify specific characteristics of firms' financial profiles of different levels of sustainability rating. Therefore, we regress daily returns of value-weighted portfolios consisting of firms of one level of sustainability ratings on the five-factor model of Fama and French (2015). The firms of different levels of sustainability ratings significantly differ in terms of idiosyncratic risk defined as the root-mean-squared error. Except for the portfolio consisting of the most sustainable stocks (rating level "A"), all portfolios consisting of sustainability levels from "B+" to "C−" (leaning toward more sustainability) exhibit lower levels of idiosyncratic risk than portfolios built with stocks rated as not sustainable (rating level "D"). To control for the differing numbers of stocks in the different portfolios, we then compute the average idiosyncratic risk for each category and find a lower idiosyncratic risk for the portfolios with higher sustainability compared to the less sustainable ones. We also document lower idiosyncratic risk for the portfolio of rating level "A" (most sustainable). Besides, we find that more sustainable portfolios exhibit lower levels of volatility (absolute risk) than less sustainable ones. Finally, we also find evidence that unsustainable portfolios (categories "D+" and "D") show negative abnormal returns.

Furthermore, the paper uncovers several substantial differences in factor exposures: Sustainable portfolios (rating level "B+") have a stronger bias toward small stocks and growth stocks than not-sustainable portfolios (rating levels "D" and "D−"). Moreover, high- and low-sustainability portfolios consist of firms with robust profitability. Finally, the pattern of the exposure to the investment factor indicates a nonlinear relationship between the level of the sustainability rating and the factor exposure. Sustainable portfolios (rating levels "A" and "B") exhibit a positive exposure to the investment risk factor, indicating a conservative investment behavior. In contrast, portfolios in the middle sustainability range (rating levels "C+," "C," "C−," and "D+") have negative coefficients.

This paper contributes to the existing literature in several ways. First, the novel dataset of Inrate sustainability ratings enables us to comprehensively add to the discussion of the effect of sustainable screening on portfolio performance, portfolio diversification, and risk factor exposures. From an asset pricing perspective, the results indicate that sustainable investments generate at least no financial cost in terms of risk-adjusted performance. Indeed, firms rated as being unsustainable show a negative performance compared to the market premium. Our empirical setting differs mainly from two streams of studies: (1) studies that investigate the relationship between corporate sustainable and financial performance (see, e.g., Friede et al. 2015; Humphrey et al. 2012) and (2) studies focusing on sustainability as an additional risk factor (see, e.g., Ziegler et al. 2007; von Arx and Ziegler 2014; Bolton and Kacperczyk 2019; Braun et al. 2019; Cao et al. 2019; Cheema-Fox et al. 2019; In et al. 2019; Zerbib 2020).

Regarding the first stream of literature, we contribute by providing first evidence on the relationship between different levels of sustainability performance and portfolio diversification as well as risk factor exposures. The validity of the empirical results of the second stream of literature is exposed to data limitations and thus faces the challenge to pass the selection criteria for a well-established additional risk factor (Lioui 2019; Fama and French 2018; Harvey et al. 2016). It is not the scope of our study to identify a valid sustainability factor. Indeed, we avoid to assume a linear relationship between sustainability and financial performance as well as diversification and risk measures by investigating different sustainably leveled groups of firms and thus mainly allow a nonlinear relationship (see, e.g., Barnett and Salomon 2006).

Second, from the portfolio manager perspective, we present a discussion on diversification characteristics and factor exposures of sustainable and unsustainable portfolios. This discussion contains valuable insights about the compatibility of sustainable themes with traditional financial risk factors, i.e., a deeper understanding of the characteristics of the financial profile of sustainably screened investment portfolios. In particular, the risk dimension and the investment style vary substantially between portfolios with different levels of sustainability ratings. The main conclusion for portfolio managers regards the fact that investments in sustainable screened portfolios do not suffer in terms of risk-adjusted performance. It also shows that firms with more sustainability tend to bear the less firm-specific risk.

The remainder of the paper is organized as follows. Section 2 presents the concepts of sustainability, sustainable investments, and a literature review. We continue with an introduction to our dataset in Sect. 3. Section 4 lays out the methodological framework, presents the results, and their implications. In Sect. 5, we show the robustness of our results and discuss some research limitations. Section 6 concludes the paper.

2 Sustainability, sustainable investments, and their characteristics

Sustainability as a term is used in a multitude of fields and by many different actors (Pufé 2014), yet with a certain degree of 'conceptual fuzziness' (Eccles and Viviers 2011). Indeed, a myriad of terms is used to describe investment practices that incorporate environmental, social, and governance factors. In this section, we precisely explain the framework in which we use the term sustainability.

2.1 The concept of sustainability

Hans Carl von Carlowitz introduced the term sustainability in his work about forestry, Sylvicultura Oeconomica, in 1713. Carlowitz, responsible for the management of forests, described with sustainability the smart exploitation of forests where the harvest of wood and plantation of trees should be undertaken together to ensure a sufficient resource of timber in the future (see Pufé 2014; Kuhlman and Farrington 2010; Staub-Bisang 2012). This principle will sound familiar to an economist since the scarcity of resources, and the allocation thereof is one of the central problems economics deal with (Kuhlman and Farrington 2010).

According to the management of scarce resources, the second half of the twentieth century was marked by the unprecedented development of human activity—referred to as 'the great acceleration' (Steffen et al. 2004; McNeill and Engelke 2014)—and its effects became increasingly apparent. Under such circumstances, the concept of sustainability gained increasing interest in the public sphere and among international bodies. Following the 1987 Brundtland report[1] aiming at giving a universal definition of sustainability and based on the United Nations' Agenda 21[2], sustainability was further defined and came to be understood as having three pillars: economic, social, and environmental (Kanning 2013; Gabriel 2014; Kuhlman and Farrington 2010; Pufé 2014). The economic dimension can be broadly defined as enabling a productive economic activity to ensure growth and development. The social dimension addresses the empowerment, inclusion, and fair treatment of all stakeholders. The environmental dimension amounts to using efficiently and preserving the resources to ensure long-term well-being (Kanning 2013; Kuhlman and Farrington 2010).

In the business world, the three pillars of sustainability are embodied by the triple bottom line approach, which underlines the three-dimensional responsibility of firms: profit, people, planet (Hahn 2013; Kuhlman and Farrington 2010). Firms usually implement the triple bottom line through their corporate social responsibility (CSR)[3] programs (Pufé 2014). Hence, sustainability in the business world goes beyond the unique goal of firms (shareholder view) to make profits and thus maximize shareholder value, which is embodied by the far-famed view of Friedman (1962) that 'the social responsibility of business is to increase its profits.' It represents an extended role given to and required from firms to consider the externalities it produces through economic activity on society and the environment (stakeholder view), which is shaped by evolving standard views on the position of the business world in society (Van Marrewijk 2003; Pufé 2014).

2.2 Sustainable investments

Parallel to the concern over sustainability issues in the business world, investors have also been considering such aspects when choosing firms to invest in (Laurence 2013). Precursors to sustainable investment find their roots in ethics, politics, and religion. With the growing concerns for sustainability in the second half of the twentieth century and because of scandals of criminal corporate practices such as the cases of Enron and Worldcom in the USA or Parmalat in Italy (2001, 2002, and 2003 respectively), an increasing number of investors have started paying attention to extra-financial criteria related to sustainability (Renneboog et al. 2008). Investment practices taking into account these criteria are usually referred to as investments applying environmental, social, and governance (ESG) criteria (Laurence 2013; Renneboog et al. 2008).

1 The United Nations, under the stewardship of the World Commission on Environment and Development, addressed the term of sustainability in the 1987 report 'Our Common Future,' which came to be known as the Brundtland Report, and gave it the following commonly accepted definition: Sustainable development is a development that meets the needs of the present without compromising the ability of future generations to meet their own needs (World Commission on Environment and Development 1987; Staub-Bisang 2012).

2 Agenda 21—'21' stands for the twenty-first century—is the action plan of the United Nations to concretize the steps necessary to implement sustainable development in various levels of society (e.g., governments, NGOs, businesses). It is the result of the first UN Conference on Environment and Development in Rio de Janeiro in 1992 (Pufé 2014).

3 Many similar terms are used to describe the efforts of firms in terms of sustainability, which include corporate sustainability, corporate responsibility, and corporate citizenship.

The tripartite structure of ESG criteria is closely related to the three pillars of sustainability in terms of the triple bottom line approach (planet, people, profits), whereby the (corporate) governance criterion has replaced the economic one since ESG criteria[4] explicitly consider extra-financial factors (United Nations Environment Programme Finance Initiative 2010). Other terms describing investments considering extra-financial information include socially responsible investment, ethical investment, responsible investment, and sustainable investment (see, e.g., Eccles and Viviers 2011; Eurosif 2018; Laurence 2013).

Several strategies exist to build sustainable investments. Negative screening, considered as the oldest strategy, consists of excluding stocks or sectors from the investable universe because they fail to meet ESG criteria. Positive screening works in the opposite direction. Instead of excluding investments, positive screening explicitly includes investments when they meet ESG standards. After the application of a positive screening approach, investors construct their portfolios based on financial objectives on the available universe. A variation of the positive screening approach is called the best-in-class strategy. In this case, the positive screening is applied separately on the sectors or industries of the universe, and the goal is to foster competition among firms in the same sectors (Eurosif 2018; Renneboog et al. 2008; Staub-Bisang 2012).

2.3 Sustainable investments' characteristics: a literature review

Investors' perception of the financial profile of sustainable investments plays an essential role in the investment decision. Value-driven investors might accept lower financial performance for sustainable investments because of the extra-financial utility they receive from such investments (Levitt and List 2007). This extra-financial utility turns into an increase in demand for sustainable investments since value-driven investors reward sustainable firms while avoiding less sustainable ones. Due to the assumed imperfection of demand curves' elasticity, stock prices will move accordingly. Furthermore, a smaller investor base bears the risk of shunned firms' stocks, limiting risk-sharing opportunities. These effects will affect expected return and thus the cost of capital, which will be lower for sustainable firms (Derwall et al. 2011; Heinkel et al. 2001; Revelli and Viviani 2015) and thus gives such firms more leeway to invest in beneficial projects.

Opponents of this view argue that such sustainable investors might not be large enough in numbers to produce the described effects (Derwall et al. 2011; Heinkel et al. 2001; Hudson 2005). Nevertheless, sustainable investments might display superior characteristics because firms with high sustainable scores display enhanced efficiency, achieve differentiation on the market, avoid regulatory pressure and costs, avoid or minimize litigation risks from stakeholders, experience improved relationships with stakeholders, and have sound management practices (Barnett and Salomon 2006; Carroll and Shabana 2010; Verheyden et al. 2016; de Carvalho Ferreira et al. 2016;

[4] The use of these broadly defined concepts to assess the extra-financial information of investments is well established (Eccles 2010), even though the scope of criteria and types of indicators used to vary and depend on the institution and the country (United Nations Principles for Responsible Investment 2017; Eurosif 2018).

Renneboog et al. 2008). High ESG scores might also be linked to good managerial quality. Moreover, the proactive approach toward risks leads to the avoidance of high scandal and crisis costs, which can translate into lower risk.

In line with the neoclassical economics view, sustainable firms generate subcommercial returns. According to this view, the role of a firm to maximize profits and the pursuit of goals related to sustainability allocate resources to other purposes, which ultimately leads to a suboptimal bottom line (Barnett and Salomon 2006; Carroll and Shabana 2010; Renneboog et al. 2008). Furthermore, sustainable screening approaches reduce the number of stocks in an investment universe. According to the modern portfolio theory, constrained universes cannot be more diversified than the conventional market portfolio (Markowitz 1952; Adler and Kritzman 2008). Thus, the application of sustainable screenings shifts the mean-variance frontier toward less favorable risk–return opportunities. In contrast, Diltz (1995) concludes that less diversification caused by sustainable investment styles has almost no effect due to the efficiency and size of the market. In particular, the incomplete information model of Merton (1987) states that portfolios of assets with concentrated information are superior to holding the market portfolio (Barnett and Salomon 2006; Bauer et al. 2005; Renneboog et al. 2008; Revelli and Viviani 2015; Sauer 1997).

A major debate in sustainable investment literature relates to the performance of such investments (Capelle-Blancard and Monjon 2012), even though very few research exists on the effect on diversification (Pizzutilo 2017). Some exceptions (e.g., Kempf and Osthoff 2007; Verheyden et al. 2016) find a higher risk-adjusted performance and little diversification constraints of portfolios screened concerning ESG best-in-class approach in a global study. Pizzutilo (2017) investigates the level of idiosyncratic risk a sustainable portfolio bears and finds constraints for diversification in the sustainable index family of the MSCI. Particularly for sustainably screened portfolios with high screening intensity, a documented loss in diversification exists (Capelle-Blancard and Monjon 2014; Humphrey and Tan 2014). On a stock level, stocks with low-sustainability assessments exhibit higher firm-specific risk, which is compensated by higher returns (Lee and Faff 2009). We add to this literature by studying diversification, performance, and investment styles characteristics of portfolios with various sustainability screens.

3 Data

Our sample contains listed firms from the MSCI ACWI, and additionally, some listed stocks not included in the MSCI ACWI from Switzerland, Argentina, Luxemburg, Morocco, and Togo. For these stocks, we have ESG ratings from Inrate, a Swiss sustainability rating agency, for a period ranging from October 2013 to May 2017. Inrate covers more than 3000 firms (both equity and debt) worldwide (including both developed and emerging markets), thereby not limiting itself to a geographic region or country. Inrate's ESG ratings consider both the products and the corporate social responsibility (CSR) of firms by evaluating a broad variety of criteria. Overall, 147 criteria determine in the subcategories of environment, labor, governance, and society (Schwegler 2018). The aggregation of these criteria results in a firm's ESG score. Each criterion and the aggregate ESG score range on a scale from "A+" to "D−."[5] The A-range denotes a firm to be "sustainable," B-range means "on the path to sustainability," C-range is a synonym of "not sustainable, but with less negative impact," and D-range includes 'not sustainable' firms (Schwegler 2018). The default updating frequency of Inrate's ratings is yearly. However, if substantial new information is available during the year, the update frequency varies between yearly, half-yearly, and quarterly.

ESG scores are not without criticism. For instance, Drempetic et al. (2019) document a size bias for rated universes. Moreover, high ratings for corporate social responsibility may result from sustainable compensation actions of firms for their high corporate social irresponsibility in certain aspects (such as tobacco, see, e.g., Kotchen and Moon 2012). Thus, measures for positive corporate social performance and negative controversies should be considered separately (Mattingly and Berman 2006).

The Inrate methodology captures such critics to some extent by considering a variety of sources and criteria. It goes beyond usual sustainability ratings by explicitly considering the product lifecycle impact on the environment and society besides to the typical CSR indicators used. The assessments for the product and the CSR dimensions are reported separately. The evaluation criteria in the product pillar are the environmental and social impact of the products and services. For CSR, Inrate considers the three classic pillars of sustainability (environment, social, and governance). Inrate's ESG ratings follow a clear modular strategy, and the layout of its method is transparent to users. The assessment of each subcategory and the final aggregate ESG score is based on a scale ranging from "A+" to "D−." Inrate rates companies without limits to a geographic region, thus enabling the analysis of global portfolios. Nevertheless, Inrate's ratings are rarely used in research on sustainable investments using ESG ratings. Thus, employing this new dataset provides robustness for existing evidence on the impact of considering extra-financial information in portfolio choice (see, e.g., Auer and Schuhmacher 2016; Hong and Kacperczyk 2009; Kempf and Osthoff 2007).

Besides ESG ratings, we also receive country and industry characteristics of the sample firms from Inrate. We augment our dataset with daily stock prices (in USD), total return indices, book-to-market ratios, market capitalization (in USD), and leverage ratios from Thomson Reuters Datastream. We retrieve the global factor returns (the market portfolio, the small-minus-big factor, high-minus-low factor, robust-minus-weak factor, and conservative-minus-aggressive factor) from the Kenneth French Data Library[6]. The risk-free rate stems from the French Data Library and is the US one-month T-bill rate.

Table 1 presents the descriptive statistics of each rating level and the entire universe. The values reported are sample period averages of the cross-sectional means (and standard deviations in parentheses). No firm in the panel achieves the highest sustainability rating category, "A+," over the sample period. Because of the small scope of the "A"-rating category, which counts on average three firms, we merge it with the

[5] A paper written by Schwegler (2018) can be found online under https://www.inrate.com/cm_data/inrate_methodology_paper_newdesign_02.pdf.
Additional information on Inrate's rating methodology can also be found under https://www.inrate.com/en/esg-impact-ratings.html.

[6] The data is from mba.tuck.dartmouth.edu/pages/faculty/ken.french.

Table 1 Descriptive statistics

	Rating category											Total
	A	A−	B+	B	B−	C+	C	C−	D+	D	D−	
Characteristics												
Firms	3	30	148	309	449	573	536	313	153	38	10	2562
Countries	2	14	27	38	42	42	43	36	30	11	7	49
Sectors	1	7	9	10	10	10	10	9	8	7	7	10
Share DM	0.95	0.87	0.85	0.91	0.86	0.83	0.81	0.80	0.76	0.85	0.81	0.84
Mkt Cap	26.45	15.16	17.04	19.42	15.95	16.16	16.34	19.55	18.34	37.84	46.00	17.43
	(8.43)	(17.29)	(21.19)	(35.08)	(26.84)	(30.88)	(36.17)	(39.98)	(21.94)	(74.45)	(64.26)	(33.75)
B/M	0.46	0.51	0.55	0.57	0.59	0.63	0.70	0.75	0.85	1.03	0.62	0.66
	(0.39)	(0.53)	(0.56)	(0.55)	(0.60)	(0.52)	(0.50)	(0.54)	(0.65)	(0.71)	(0.47)	(0.56)
Subcategory scores												
Environment	11.54	10.31	9.13	8.08	7.04	5.84	4.82	3.72	2.59	1.33	5.11	5.78
	(0.36)	(0.80)	(0.98)	(1.05)	(1.15)	(1.18)	(1.13)	(1.19)	(1.07)	(0.57)	(3.23)	(2.17)
Labor	8.30	8.98	8.59	7.69	6.15	5.49	4.48	3.89	2.83	1.91	2.57	5.43
	(1.26)	(2.07)	(1.63)	(1.78)	(1.97)	(1.94)	(1.78)	(2.11)	(1.70)	(1.18)	(2.05)	(2.46)
Governance	9.02	10.23	10.10	9.60	8.97	8.66	7.85	7.89	8.12	7.97	7.93	8.59
	(0.71)	(1.65)	(1.83)	(1.99)	(2.06)	(2.13)	(2.36)	(2.73)	(2.66)	(2.38)	(2.75)	(2.37)
Society	9.81	9.09	8.23	7.49	6.94	6.25	5.41	4.61	3.91	3.36	4.17	6.07
	(0.33)	(1.33)	(1.40)	(1.37)	(1.41)	(1.40)	(1.30)	(1.32)	(1.45)	(1.53)	(1.47)	(1.84)

Table 1 continued

	Rating category											Total
	A	A−	B+	B	B−	C+	C	C−	D+	D	D−	
Probability of rating category change in %												
≥ 2	n/a	0.0	0.2	0.7	1.1	1.3	2.1	2.3	3.3	2.0	1.6	1.3
+1	0.0	2.3	3.7	6.9	7.4	8.4	10.2	11.4	9.9	12.9	3.1	6.9
=	90.0	87.9	87.4	87.1	86.7	86.2	84.3	84.1	84.9	85.2	95.3	87.2
− 1	10.0	9.8	7.3	4.8	4.3	4.0	3.1	2.0	1.9	0.0	n/a	4.3
≤−2	0.0	0.0	1.4	0.5	0.5	0.2	0.2	0.1	0.1	n/a	n/a	0.3

The table presents summary statistics of the cross section of stocks in each rating category (A to D−) over the period from October 2013 to May 2017. The displayed numbers are average values; the numbers in parenthesis are the average standard deviations. In detail, we report the average number of firms, the average number of countries, and the average number of sectors. The share of DM represents the share of market capitalization from developed markets (DM), scaled to 1 (unit interval). The remaining part comes from emerging markets (EM). Market capitalization is in billions USD, and the row B/M contains average book-to-market ratios. The subcategory scores part presents the scores of the subcategories environment, labor, society, and governance. The panel "Probability of rating category change in %" contains the probability of rating migration in a rating update. The rating update deadlines are 30.09.2013, 31.03.2014, 31.03.2015, 30.09.2015, 31.03.2016, 31.12.2016, and 31.03.2017

"A−" category for subsequent analysis and name this category "A." This gives an average of 33 firms in this new "A" category and a total number of ten categories (A, B+, B, B−, C+, C, C−, D+, D, D−). The number of firms is highest for the middle-range ratings (highest in "B−," "C+," "C") and decreases from the center to the sides. The share of firms from developed markets is highest for the most sustainable firms and decreases with the sustainability level. The market value remains relatively constant over the categories except for the "D" and "D−" ratings that are dominated by large capitalization. On subcategory level, the scores of governance are substantially higher than the scores of the three other subcategories, especially for low-rated ("D−" to "B+") firms. We also report on the migration probabilities between rating levels in Table 1. The probability of improving the firm rating is higher than for a downgrade for all rating categories between "B" and "D−." The highest rating category "A" seems to be quite resilient to considerable changes as no firm has moved to a rating lower than "A−" from one update to another.

4 Empirical design and results

4.1 Portfolio construction

We start our analysis by mimicking the usual negative screening and best-in-class approaches to generate sustainable investments. For the negative screening strategy, we construct the investment universe by excluding firms with overall ESG ratings lower than a certain threshold. In particular, we create ten investment universes, the first consisting of all firms (with ratings from "A" to "D−"). With increasing screening intensity, we raise the sustainable requirements by one for each subsequent investment universe and exclude the respectively lowest level of ESG ratings consecutively. Thus, the second investment universe consists of all firms with ESG rating from "A" to "D." The last investment universe comprises all firms with ESG rating "A." The portfolio for each level of screening intensity is a value-weighted portfolio of all firms included in the respective investment subuniverse (following Auer and Schuhmacher 2016; Kempf and Osthoff 2007). Inrate published rating updates on 30.09.2013, 31.03.2014, 31.03.2015, 30.09.2015, 31.03.2016, 31.12.2016, and 31.03.2017, and we update our screening results and rebalance the portfolios accordingly.

For our second screening approach, the best-in-class strategy, we select the best x% firms in terms of sustainability in each sector. Due to the low variation among the ESG rating, we employ the Inrate's product assessment and social impact rating as the sustainability assessment to determine the best x%. This score measures the impact on sustainability and society of the firms' products and services throughout their life cycles (Schwegler 2018). We choose to have ten steps of 10% each. The first universe thus selects the best 100% firms of each sector, i.e., all stocks. The second broadest universe contains the best 90% of each sector. We generate value-weighted portfolios of the resulting investment subuniverses that we rebalance at the screening update dates.

4.2 Statistical inference using block bootstrap

We apply a bootstrap approach to test for statistical inference of our results. A vast strand of literature documents the improved accuracy of bootstrapping compared to using asymptotic normality (see, e.g., Auer and Schuhmacher 2016; Lahiri 2003). The statistics under scrutiny in this paper are, among others, the coefficients and the error term of the five-factor asset pricing model suggested by Fama and French (2015). To be able to conduct the regression on each of the bootstrapped samples, the returns of the portfolios and the different factors are bootstrapped in pairs to preserve the relationship (Davison and Kuonen 2002; Fox 2015; Lai and Xing 2008). For this purpose and building upon Dichtl et al. (2014) and Ledoit and Wolf (2008), we employ the stationary block bootstrap of Politis and Romano (1994).

This method, which preserves most of the time-series properties (Dichtl et al. 2014), randomly picks blocks of returns from the historical time series of returns (with replacement) and combines them into a new bootstrapped sample of the same length as the original sample. The size of the block length is randomly defined from a geometrical distribution with an expected value of 2 following Dichtl et al. (2014); Patton et al. (2009); Politis and White (2004). Another feature of the bootstrapping method applied in the paper is its circularity, meaning that the blocks encompassing returns at the end of the historical sample will continue at the beginning of the historical sample. This approach ensures that the returns appear in an equal-distributed fashion in the bootstrapped samples. According to Ledoit and Wolf (2008), our choice of 5000 for the number of simulations is sufficient for practical purposes.

4.3 Main results on portfolio characteristics

Table 2 contains the portfolio characteristics of the negative screening portfolios, the best-in-class portfolios, and the unrestricted investment universe acting as a benchmark portfolio. The first column shows the annualized mean excess return $\bar{\mu} = \left(\sqrt[T]{(1+R_1)\ldots(1+R_T)}\right)^m - 1$ where R_i represent the excess returns in period i, T is the number of periods i in the estimation window, and m is the factor in annualizing the estimates (e.g., for monthly returns $m = 12$). The second column presents the annualized standard deviation of excess returns $\bar{\sigma}$, and the third column reports the Sharpe ratio $SR = \bar{\mu}/\bar{\sigma}$. For the screened portfolios, the annualized excess return increases with the screening intensity. The standard deviation of the excess returns is rather stable at the level of the one of the unrestricted universe, except for the portfolio comprising of "A"-rated firms (12.82). Moreover, negative-screened portfolios tend to have higher Sharpe ratios than the unrestricted universe. However, based on bootstrapped statistical inference, only portfolios with screening intensities "A to D+" to "A to C+" have Sharpe ratios that statistically are different from the unrestricted universe. For the best-in-class portfolios, we find Sharpe ratios very close to that of the unrestricted universe.

One caveat of actively managed portfolios such as sustainable investments is a comparatively high level of turnover. To assess the amount of trading required for

Table 2 Characteristics of screened and best-in-class portfolios

Portfolio	Return (%)	Std (%)	Sharpe	Turnover	Abn. ret.(%)	Idio. risk	Stocks	Countries	Sectors	% DM
Unrestricted investment universe										
A to D−	7.38	10.53	0.70	0.10	−0.49	0.0015	2535	49	10.00	0.68
Screened portfolios (increasing sustainability)										
A to D	7.38	10.56	0.70	0.10	−0.47	0.0015	2526	49	10.00	0.68
A to D+	7.75	10.51	0.74*	0.10	−0.07**	0.0015	2488	49	10.00	0.68
A to C−	8.29	10.38	0.80**	0.11	0.51***	0.0015	2336	48	10.00	0.69
A to C	8.56	10.39	0.82*	0.15	0.65**	0.0016	2026	48	10.00	0.71
A to C+	8.63	10.55	0.82*	0.20	0.99*	0.0017	1494	47	10.00	0.74
A to B−	8.96	10.61	0.84	0.23	1.20	0.0019	928	44	9.80	0.77
A to B	9.88	10.78	0.92	0.29	1.41	0.0023	485	40	9.60	0.79
A to B+	8.63	10.81	0.80	0.45	0.52	0.0028	179	31	9.00	0.80
A	12.04	12.82	0.94	0.56	5.70*	0.0047	33	14	7.13	0.81

Table 2 continued

Portfolio	Return (%)	Std (%)	Sharpe	Turnover	Abn. ret.(%)	Idio. risk	Stocks	Countries	Sectors	% DM
Best-in-class portfolios (increasing sustainability)										
Best 90%	7.18	10.61	0.68	0.16	−0.50	0.0015	2289	48	10.00	0.70
Best 80%	7.39	10.62	0.70	0.20	−0.51	0.0015	2036	48	10.00	0.71
Best 70%	7.22	10.65	0.68	0.23	−0.80	0.0017	1785	48	10.00	0.73
Best 60%	7.36	10.74	0.68	0.24	−0.70	0.0017	1529	48	10.00	0.74
Best 50%	7.54	10.72	0.70	0.25	−0.69	0.0017	1274	47	10.00	0.75
Best 40%	7.68	10.80	0.71	0.30	−0.70	0.0019	1022	46	10.00	0.76
Best 30%	7.66	11.09	0.69	0.34	−0.81	0.0022	769	45	10.00	0.77
Best 20%	7.91	11.33	0.70	0.37	−0.46	0.0025	515	43	10.00	0.78
Best 10%	7.75	11.86	0.65	0.48	0.12	0.0030	265	37	10.00	0.81

This table reports upon portfolio characteristics over the period from October 2013 to May 2017. Return refers to the average annualized excess return, Std to the standard deviation of the annualized excess return, Sharpe to the Sharpe ratio, Turnover to the annualized turnover rate, Abn. ret. to the abnormal portfolio return after controlling for Fama and French (2015) risk factors, Idio. risk to the idiosyncratic risk measured by the root-mean-squared error (RMSE) from the five-factor model, Stocks to the number of stocks in the portfolio, Countries to the number of countries in the portfolio, Sectors to the number of sectors in the portfolio, and %DM to the ratio of firms from developed market in the portfolio. Statistical tests for differences between the unrestricted investment universe and the respective selected portfolios were performed on the risk-adjusted performance measures (i.e., Sharpe ratio and the abnormal return) using the bootstrap approach referred to in Sect. 4.2

* indicates statistics that are statistically significant at the 10% level, ** at 5%, and *** at 1%

our investment strategies, we compute the average of the turnover ratio (τ) at each rebalancing date following DeMiguel et al. (2009):

$$\tau = \sum_{i=1}^{N} (|w_{i,t+1} - w_{i,t^+}|) \tag{1}$$

$w_{i,t+1}$ represents the portfolio weight for each asset i before the rebalancing at the end of the rebalancing period t and w_{i,t^+} is the weight after the rebalancing, but before $t + 1$ has started. We define turnover as the sum of the absolute value of the rebalancing movements across all assets. Column "Turnover" of Table 2 contains the annualized values of the turnover ratio. Turnover rates increase with higher levels of sustainability intensity. In absolute values, turnover rates are higher than the ones reported in DeMiguel et al. (2009) for the $1/N$ portfolio, yet for the actively managed portfolios reasonably small compared to the results of DeMiguel et al. (2009). Although we analyze value-weighted portfolios, turnover rates are different from zero since the constituents list of the respective investment universe varies with rating up- and downgrades.

Column "Abn. ret. (%)" of Table 2 contains the abnormal return. The abnormal return is calculated using the five-factor asset pricing model of Fama and French (2015):

$$\begin{aligned} R_{i,t} - R_{f,t} = a_i + b_i(R_{M,t} - R_{f,t}) + s_i R_{\text{SMB},t} \\ + h_i R_{\text{HML},t} + r_i R_{\text{RMW},t} + c_i R_{\text{CMA},t} + e_{i,t} \end{aligned} \tag{2}$$

a_i represents the abnormal return of portfolio i. $(R_{M,t} - R_{f,t})$ represents the excess return of the market index, R_{SMB} is the return of the small-minus-big portfolio, a zero-investment portfolio long in stocks with small capitalization and short in stocks with large capitalization, and R_{HML} is the return of the high-minus-low portfolio, i.e., high book-to-market value against low book-to-market value. The remaining two risk factors are R_{RMW} and R_{CMA}, which stand for robust-minus-weak and conservative-minus-aggressive, respectively. The former represents the excess return of robust profitability stocks against weak profitability stocks (profitability factor). The latter factor is long in firms with low-investment activities and short in those with high-investment activities.

While screened portfolios with high sustainability intensity show positive abnormal returns, best-in-class portfolios (except the one with the best 10%) generate negative abnormal returns. The abnormal return increases with the screening intensity. Compared with the unrestricted universe, only moderately screened portfolios ("A to D+" to "A to C") exhibit significantly (at 5% significance level) higher abnormal returns. Best-in-class portfolios exhibit no abnormal returns different from the unrestricted universe. Thus, the results provide supporting evidence for the no-linkage hypothesis, in general. This hypothesis argues that there is no significant difference between risk-adjusted performance (alpha) of portfolios exhibiting varying sustainability levels (Lee et al. 2013). The fact that portfolios consisting of sustainable firms (positive screening) earn higher returns is in line with Statman and Glushkov (2009). They find higher expected excess returns of sustainable portfolios compared to conventional investments for a US sample.

Furthermore, Table 2, Column 7, reports upon the idiosyncratic risk of the different sustainability levels. We follow Lee and Faff (2009) and use the root-mean-squared error (RMSE) from the five-factor model (Eq. 2) to determine idiosyncratic risk. Idiosyncratic risk of the portfolio is the risk which is not compensated by financial return and indicates the level of diversification of a portfolio. Screened and best-in-class portfolios (including more than 1000 firms) exhibit a similar level of the idiosyncratic risk to the unrestricted universe. Idiosyncratic risk increases for screening levels with a lower number of stocks, i.e., for higher sustainable rating intensity.

In summary, the results of Table 2 show that screened portfolios with mildly restrictive screening intensity show positive performance and diversification characteristics when compared to the unrestricted universe. These portfolios exhibit a similar exposure to countries and sectors. In particular, portfolios with a negative screening level of "A to C+" comprise firms of all sectors and almost all countries (47 out of 49).[7] Increasing the sustainable screening intensity from "A to C+" causes a rapid decrease in the number of countries. This effect can be observed in the increasing idiosyncratic risk and general risk. The last column of Table 2 contains the share of stocks from developed markets (DM). This share increases with higher sustainability screening intensity.

The main characteristic of best-in-class portfolios—due to their construction—is the occurrence of all sectors regardless of the screening intensity. We find lower variability in returns between the different universes compared to screened portfolios. Moreover, for best-in-class portfolios, we find no positive relationship between the Sharpe ratio and screening intensity. The best performing portfolio is the one containing the best 40% of stocks with 0.71 and decreases onward with selection intensity. In terms of diversification, the number of countries and stocks differs from the screened portfolios. The portfolios consisting of the best 10% of stocks contain, on average, 265 stocks from 37 countries. The idiosyncratic risk increases together with selection intensity but remains low. The portfolios comprising at least the best 40% of firms have an idiosyncratic risk similar to that of the entire universe. From a risk-taking perspective, the portfolio-specific risk an investor accepts when investing sustainably (until the 40% threshold) is similar to the one of the unrestricted universe. Analog to the screened portfolios, we detect a positive relationship between the screening intensity and both the share of stocks from developed markets and portfolio turnover.

The results of the best-in-class portfolios also support the no-linkage hypothesis. Since screened and best-in-class portfolios exhibit different numbers of stocks in certain levels of screening intensity, a direct comparison has to be applied with caution. The best-in-class approach appears to generate portfolios with a higher level of diversification (in terms of idiosyncratic risk) compared with the screened portfolios on a first glance. However, as soon as we relate the performance and diversification measures of best-in-class portfolios to screened portfolios with a similar number of stocks, screened portfolios exhibit higher returns, lower standard deviations, higher Sharpe ratios, and lower turnover. The results indicate the financial characteristics of

[7] The unrestricted universe includes firms from 49 countries.

sustainably screened portfolios in general. The approach is aligned with the practical applications of investors. Nevertheless, this approach does not allow us to conclude the characteristics of specific sustainability rating levels. Therefore, we continue with an analysis of the differences between firms of separate sustainability levels in the next section.

4.4 Discussion on portfolio characteristics of separate sustainability levels

Table 3 reports upon the annualized value-weighted portfolio characteristics derived from the historical realizations and the coefficients estimated in Fama and French (2015) five-factor models. Additionally, we illustrate the results of the bootstrap simulations in Fig. 1. We bootstrapped confidence intervals for all statistics. Due to the nature of bootstrapping, these intervals are wider than the ones of the regression model, and thus statistical inference is more robust.

The groups of firms with the lowest average return and the highest risk are sustainability levels "D" and "D+." The rating categories with the highest number of stocks (between "B" and "C−") have the lowest levels of volatility and idiosyncratic risk because of the diversification gains from the high number of firms. The turnover ratio, which informs about how much of the portfolio has been traded over a year, is highest for the ratings "A" and "B+" and lowest for the "D−" rating.

Our measure for idiosyncratic risk is the root-mean-squared error (RMSE) of the Fama and French (2015) model. The reduction of idiosyncratic risk for firms with ESG ratings confirms a risk-reducing effect of high sustainability (Derwall et al. 2011; Humphrey and Tan 2014; Lee and Faff 2009; Utz 2018). As expected, the portfolios with a higher number of stocks have a better diversification resulting in lower idiosyncratic risk. The diversification benefits stemming from larger portfolios likely outweigh the gains from the high sustainability ratings at this point ("D+" has 152 stocks on average, whereas "A" has 33). When comparing portfolios with a similar number of stocks, for instance, "B+" (146 stocks) with "D+" (152 stocks), the idiosyncratic risk is significantly higher for less sustainable portfolios.

To control for the different portfolio sizes ("D−," the least sustainable category, has an average of ten firms, while category "C+" has 566 firms on average) and the associated gains in diversification that arise, we compute the idiosyncratic risk for each firm and each subperiod (i.e., in-between rating updates) and then compute the average, value-weighted RMSE for each category. So doing, we can observe the average level of idiosyncratic risk for each category independent of the size of the portfolio. Again, the results show clear indications in favor of the risk-reducing effect of high sustainability. Firms from the most sustainable category ("A") have the second-lowest average idiosyncratic risk of the ten categories; firms from the second-most sustainable category ("B+") exhibit the third-lowest average idiosyncratic risk. Category "D−" has the lowest level. This finding is mainly biased by one large US retail firm that covers almost half of the portfolio weight of this category (44–54% depending on the time interval) and has a comparably low idiosyncratic risk. When excluding this firm, the average RMSE among category "D−" firms goes up to 0.0142, which is the second-highest RMSE of all categories. This additional analysis thus provides us with

Table 3 Separate rating category, value-weighted portfolio results

	A	B+	B	B−	C+	C	C−	D+	D	D−
Portfolio characteristics (ann.)										
Return (%)	12.04	8.03	10.56	7.71	8.00	8.37	6.86	0.84	−2.15	7.48
Std (pf.) (%)	12.82	10.76	10.92	10.61	10.66	10.28	10.81	14.60	15.90	12.18
Sharpe ratio	0.94	0.75	0.97	0.73	0.75	0.81	0.64	0.06	−0.14	0.61
Stocks	33	146	305	443	566	532	310	152	38	10
Turnover ratio	0.56	0.53	0.43	0.47	0.46	0.48	0.38	0.35	0.27	0.26
RMSE (pf.)	0.0047	0.0029	0.0022	0.0019	0.0019	0.0020	0.0021	0.0040	0.0051	0.0048
RMSE (st.)	0.0126	0.0134	0.0135	0.0138	0.0138	0.0137	0.0140	0.0168	0.0135	0.0118
Std (st.) (%)	21.84	23.26	23.86	24.36	24.32	24.13	24.39	28.89	23.65	21.06

Table 3 continued

	A	B+	B	B−	C+	C	C−	D+	D	D−
Results from five-factor model										
Abn. return (% ann.)	5.70	−1.04	2.01	1.10	0.52	0.45	0.38	−6.99**	−11.48***	3.73
t-statistics	1.45	−0.45	1.09	0.71	0.33	0.27	0.22	−2.25	−2.95	0.95
$R_{M,t} - R_{f,t}$	0.97	0.97*	1.00	0.95***	0.95***	0.89***	0.88***	1.09***	0.95*	0.72***
t-statistics	−1.15	−1.95	0.43	−3.84	−4.09	−9.15	−9.09	4.09	−1.76	−9.95
SMB	−0.06	0.18***	0.00	−0.05**	−0.06***	−0.09***	−0.11***	−0.03	−0.51***	−0.25***
t-statistics	−1.19	5.58	0.02	−2.32	−2.77	−4.23	−4.47	−0.75	−9.00	−4.75
HML	−0.25***	−0.05	−0.20***	−0.12***	0.02	0.10***	0.33***	0.88***	0.81***	0.00
t-statistics	−3.67	−1.21	−6.15	−4.31	0.64	3.42	10.51	15.31	10.95	−0.04
RMW	0.13***	0.38***	0.21***	0.08**	0.12***	0.13***	0.02	0.04	0.11	0.48***
t-statistics	1.6	7.51	5.28	2.44	3.57	3.6	0.6	0.56	1.24	5.71
CMA	0.41***	0.00	0.16***	−0.05	−0.14***	−0.19***	−0.32***	−0.26***	0.15	0.07
t-statistics	3.82	−0.01	3.14	−1.24	−3.08	−4.06	−6.49	−2.85	1.31	0.63
R^2 (adj.)	0.64	0.82	0.89	0.92	0.92	0.90	0.90	0.81	0.71	0.52

This table presents key annualized value-weighted portfolio characteristics of the portfolios and the regression results of the five-factor model over the period from October 2013 to May 2017. Std. stands for standard deviation. The robust regression coefficients are reported together with their respective t-statistics. The null hypothesis per default is $H_0 = 0$. The only exception is the variable $R_{M,t} - R_{f,t}$, for which the null hypothesis is $H_0 = 1$. RMSE stands for root-mean-squared error and is a proxy for the idiosyncratic risk (see, e.g., Lee and Faff 2009). RMSE (pf.) is the category portfolio's idiosyncratic risk, while RMSE (stocks) is the value-weighted average of all firms' RMSEs for the respective rating categories. Std (stocks) is the value-weighted average of all firms' daily annualized volatility in % for the respective rating categories. The rebalancing is undertaken after every rating update

* indicates coefficients that are statistically significant at the 10% level, ** at 5%, and *** at 1%

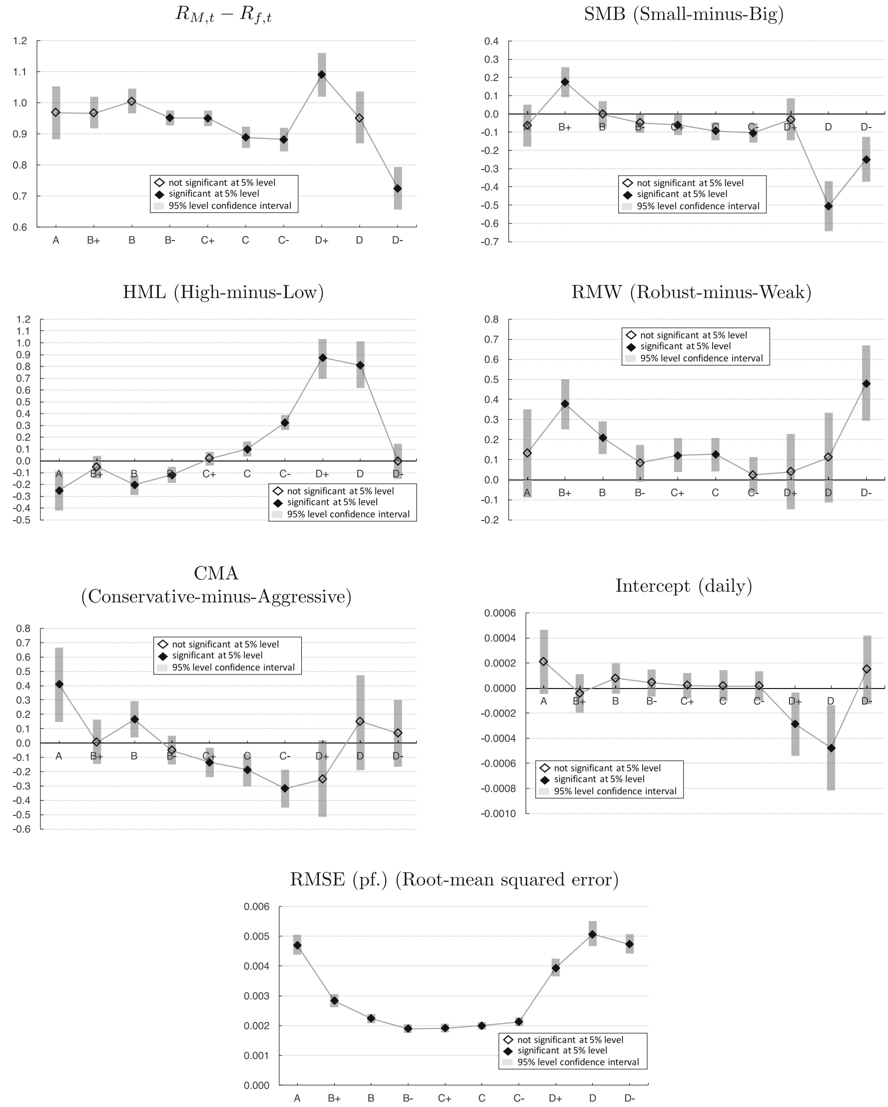

Fig. 1 Value-weighted portfolios bootstrap coefficient estimates: means and confidence intervals

additional evidence that more sustainable firms are less prone to idiosyncratic risk than categories made of less sustainable firms. Performing two-sample *t*-tests on the RMSE levels for the different categories, the hypothesis that the RMSE of the different categories come from a normal distribution with equal means can be rejected a the 5% level for the majority of pairs.

We also find that the most sustainable portfolio has volatility much lower than that of the "D" portfolio, while both consist of a comparable number of stocks. To study the absolute risk levels, we proceed with the same above-mentioned method to test for the absolute level of risk (volatility) for the different portfolios on firm-level and compute the average, value-weighted, annualized daily volatility for each category to observe the absolute risk characteristics independent of the number of stocks in the portfolio. Again, we find that the portfolio with the highest sustainability ("A") displays the lowest volatility after the least sustainable one. The reason is the same as before: The least sustainable is dominated by a large US retail firm, which exhibits low volatility. Without this firm, the volatility of category "D−" is substantially higher than that of category "A." The results provide the same insight again: The more sustainable categories ("A" and "B+") have levels of risk—measured as volatility—lower than that of less sustainable portfolios ("D+" and "D" for instance). Again, those results are statistically significant at the 5% level for most pairs when performing the two-sample t-test for the different combinations.

Moreover, we find clear evidence that firms of rating level "D+" are the only group that has a higher exposure to systematic risk than the market portfolio, i.e., the market beta (1.09) is significantly higher than one. The coefficients of the SMB (small-minus-big) risk factor indicate that the lower the sustainability rating is, the larger the firms are. Although value-weighted portfolios favor large stocks by definition, the "B+" portfolio has a significant bias toward small stocks, which is in line with the findings of Bauer et al. (2005) on ethical funds. In general, the majority of the different rating levels comprise large caps due to the significantly negative coefficients of the size factor in six out of ten levels. The factor loadings on the HML factor are negative for high sustainability ratings and positive for low sustainability ratings. Thus, growth stocks tend to have higher sustainability ratings, while the portfolios of low sustainability rating levels have a bias toward value stocks. Firms with high sustainability ratings often offer products and services with an orientation to the future supporting sustainable development. This finding is consistent with the growth bias of sustainably top-rated portfolios. The profitability factor is nonnegative for all the rating levels. However, portfolios with high and low sustainability ratings show positive significant coefficients indicating a bias to robust firms.

Finally, the investment risk factor indicates a nonlinear relationship between the rating level and the factor exposure. The portfolio of rating level "B" shows a positive significant coefficient indicating conservative investment behavior. In contrast, portfolios of rating levels "C+," "C," "C−," and "D+" show significant negative coefficients indicating aggressive investment behavior. "D" and "D−" portfolios have insignificantly positive coefficients. A possible explanation regards the fact that firms with a high level of sustainability have more future-oriented management, and managers are less likely to deviate from optimal risk-taking levels (Harjoto and Laksmana 2018).

In summary, we find that investors have no cost to invest sustainably in terms of abnormal returns. This finding supports the no-linkage hypothesis, yet does not support the curvilinear relationship between financial and sustainability performance (Barnett and Salomon 2006), for instance. Nevertheless, our study is no contradiction to such earlier findings since our sample selection (stocks vs. screened SRI funds), our observation period (2013–2017 vs. 1972–2000), and the calculation of abnormal returns (Fama and French (2015) model vs. capital asset pricing model) substantially differ from earlier studies. Indeed, our results show that portfolios built upon the unsustainable firms of rating levels “D+” (at the 5% level) and “D” (at the 1% level) show below market-rate returns.

5 Robustness checks and research limitations

5.1 Robustness checks

Since specific choices in the data selection and the methodological approach may influence the results, we apply several robustness checks. First, we change the rebalancing frequency of our portfolios since the rebalancing frequency may impact the performance of a strategy. The frequency used in our analysis was to rebalance after an update in the Inrate database. The periodicity of these updates is not equidistant ranging from 3 months to 6, 9, and 12 months. For this robustness check, we apply monthly rebalancing for the portfolio construction. The rebalancing frequency does not change the conclusions of the analysis above, and all values keep the same qualitative values (signs).

As a second robustness check, we base our sample on weekly returns instead of daily returns. This approach may alleviate the effects of different opening hours. We run all analyses based on weekly returns and find that this additional robustness test does not change the results qualitatively.

Finally, the construction of the chosen portfolio might also influence the results. To account for this, we apply equal-weighted portfolios instead of value-weighted portfolios and perform the same tests. The overall results do not differ substantially. Comparing the results of the two methods, we find that value-weighted construction produces higher performance and risk measures for single-category portfolios at the extremes. At the same time, the equal-weighted portfolios do so for portfolios in the middle range. We also find that equal-weighted portfolios generally carry a more substantial share of idiosyncratic risk. In particular, equal weighting of the firms instead of value weighting for the computation of the RMSE makes the idiosyncratic risk for the most sustainable category (“A”) higher. However, we still observe a clear pattern of higher sustainability categories (“A,” “B+,” “B,” “B−”) having a lower risk than the less sustainable ones (“C−,” “D+,” “D,” “D−”). This confirms our finding that high sustainability lowers idiosyncratic risk. Looking at the volatility on stock’s level for each category [Std (st.) in Table 4], the alternative weighting confirms that categories with higher sustainability (“A,” “B+,” “B,” “B−”) have a lower absolute risk (measured as volatility) than less sustainable categories (“C−,” “D+,” “D”).

Furthermore, equal-weighted portfolios are more positively related to the SMB factor. Another deviation in factor exposure between value-weighted and equal-weighted portfolios occurs with the investment factor. It becomes significant and positive for low-rated portfolios, hinting at the fact that smaller unsustainable firms tend to invest more aggressively than larger unsustainable ones.

Table 4 Separate rating category, equal-weighted portfolio results

	A	B+	B	B−	C+	C	C−	D+	D	D−
Portfolio characteristics (ann.)										
Return (%)	11.29	7.59	10.79	9.53	7.16	6.67	5.13	1.69	−8.27	4.72
Std (pf.) (%)	13.64	11.55	11.09	10.59	10.26	10.24	10.99	14.73	16.43	13.17
Sharpe ratio	0.83	0.66	0.97	0.90	0.70	0.65	0.47	0.12	−0.50	0.36
Stocks	33	146	305	443	566	532	310	152	38	10
Turnover ratio	0.59	0.51	0.49	0.50	0.51	0.51	0.52	0.51	0.49	0.30
RMSE (pf.)	0.0055	0.0033	0.0030	0.0027	0.0025	0.0031	0.0034	0.0048	0.0062	0.0054
RMSE (st.)	0.0161	0.0152	0.0155	0.0158	0.0160	0.0163	0.0173	0.0200	0.0203	0.0154
Std (st.) (%)	29.38	26.97	27.55	28.58	29.22	29.52	30.88	35.49	36.30	27.45

Table 4 continued

	A	B+	B	B−	C+	C	C−	D+	D	D−
Results from five-factor model										
Abn. return (% ann.)	2.47	0.45	1.83	1.17	−0.87	−0.42	−2.83	−5.24	−15.52***	−3.37
t-statistics	0.55	0.17	0.75	0.53	−0.42	−0.17	−1.06	−1.37	−3.35	−0.78
$R_{M,t} - R_{f,t}$	1.03	1.00	1.01	0.97*	0.92***	0.86***	0.91***	1.02	0.93**	0.88***
t-statistics	0.99	0.24	0.80	−1.85	−5.58	−7.55	−4.60	0.61	−1.98	−3.72
SMB	0.15**	0.32***	0.41***	0.37***	0.33***	0.40***	0.47***	0.34***	0.28***	0.32***
t-statistics	2.40	8.90	12.51	12.34	11.75	11.59	12.71	6.35	4.13	5.23
HML	−0.26***	0.14***	0.16***	0.12***	0.26***	0.41***	0.60***	0.92***	1.16***	0.40***
t-statistics	−3.26	2.87	3.78	3.17	7.05	9.06	12.41	13.23	12.96	5.04
RMW	0.27**	0.42***	0.45***	0.30***	0.31***	0.38***	0.44***	0.08	0.17	0.40***
t-statistics	2.79	7.25	8.58	6.33	6.97	6.98	7.40	0.98	1.51	4.12
CMA	0.50***	0.07	0.03	−0.09	−0.20***	−0.35***	−0.44***	−0.49***	−0.86***	−0.52***
t-statistics	3.96	0.92	0.51	−1.42	−3.51	−5.00	−5.75	−4.44	−6.05	−4.20
R^2 (adj.)	0.58	0.79	0.81	0.84	0.84	0.77	0.76	0.71	0.58	0.53

This table presents key annualized equal-weighted portfolio characteristics of the portfolios as well as the regression results of the five-factor model over the period from October 2013 to May 2017. Std. stands for standard deviation. The robust regression coefficients are reported together with their respective *t*-statistics. The null hypothesis per default is $H_0 = 0$. The only exception is the variable $R_{M,t} - R_{f,t}$, for which the null hypothesis is $H_0 = 1$. RMSE stands for root-mean-squared error and is a proxy for the idiosyncratic risk (see, e.g., Lee and Faff 2009). RMSE (pf.) is the category portfolio's idiosyncratic risk, while RMSE (stocks) is the equal-weighted average of all firms' RMSEs for the respective rating categories. Std. (stocks) is the equal-weighted average of all firms' daily annualized volatility in % for the respective rating categories. The rebalancing is undertaken after every rating update

* indicates coefficients that are statistically significant at the 10% significance level, ** refers to the significance level of 5%, and *** to the significance level of 1%

5.2 Research limitations

One limitation of this study is the different country coverage of global factor returns from the Kenneth French Data Library, including 23 developed markets[8], and the Inrate ratings, including firms from 49 countries. Fama and French (2015) factors focus on developed markets. Moreover, local factors are better at explaining asset returns (Fama and French 2012; Griffin 2002). Although our analysis might fail to optimally explain the returns of international portfolios composed of stocks from developed and emerging countries, the global factors might at least partially capture the region-specific effects for the majority of the stocks being from developed markets. A second limitation is the relatively short observation period of approximately four years between October 2013 to May 2017. The observation period does neither include systematic tail events such as an economic crisis nor covers a full economic cycle. However, it does cover more turbulent market phases (Q4 2015 to Q1 2016).

6 Conclusion

This paper investigates the performance, risk, and diversification characteristics of sustainable investments. Notably, we seek to answer the question of whether portfolios with varying sustainability levels display different investment styles and whether more sustainable portfolios carry less unsystematic risk than less sustainable portfolios. Using a novel sustainability rating dataset from Inrate, we build value-weighted portfolios according to both the screening approach and the best-in-class approach for the period ranging from October 2013 to May 2017. The results on portfolio level indicate that portfolios with high levels of sustainability suffer from increasing idiosyncratic risk when the screening intensity increases, which is due to the reduction in the number of available stocks; the Sharpe ratio and the return increase, however. The best-in-class approach to sustainability implementation provides higher levels of diversification (compared to negative screening) even for portfolios with high screening intensity since more firms and different sectors are considered in the final investment decision. The abnormal return from the five-factor model of Fama and French (2015) is close to the ones of the unrestricted universe for any of the portfolios. Thus, these findings are in favor of the no-linkage hypothesis between financial performance and the level of sustainability of a portfolio. Considering the different sustainability rating categories separately, we find that the firms of the more sustainable portfolios have significantly lower levels of idiosyncratic and absolute risk. This finding supports the risk-reducing characteristics of high sustainably-rated firms. Moreover, we find higher exposure toward the high-minus-low and to the conservative-minus-aggressive factor for more sustainable portfolios. Finally, we document that low-rated portfolios underperform higher rated ones in terms of abnormal return.

[8] Australia, Austria, Belgium, Canada, Germany, Denmark, Finland, France, Great Britain, Greece, Hong Kong, Ireland, Italy, Japan, the Netherlands, Norway, New Zealand, Portugal, Singapore, Spain, Sweden, Switzerland, USA.

Acknowledgements We thank Markus Schmid (the editor) and two anonymous referees for their comments that helped us to improve the article. Furthermore, we thank Inrate AG for kindly providing us with the important ESG dataset for the paper.

References

Adler, T., Kritzman, M.: The cost of socially responsible investing. J. Portf. Manag. **35**(1), 52–56 (2008)
Areal, N., Cortez, M.C., Silva, F.: The conditional performance of US mutual funds over different market regimes: do different types of ethical screens matter? Fin. Mark. Portf. Manag. **27**(4), 397–429 (2013)
Auer, B., Schuhmacher, F.: Do socially (ir)responsible investments pay? New evidence from international ESG data. Q. Rev. Econ. Finance **59**, 51–62 (2016)
Barnett, M., Salomon, R.: Beyond dichotomy: the curvilinear relationship between social responsibility and financial performance. Strateg. Manag. J. **27**(11), 1101–1122 (2006)
Bauer, R., Koedijk, K., Otten, R.: International evidence on ethical mutual fund performance and investment style. J. Bank. Finance **29**, 1751–1767 (2005)
Bolton, P., Kacperczyk, M.: Do investors care about risk? SSRN Working Paper (2019)
Bouslah, K., Kryzanowski, L., M'Zali, B.: The impact of the dimensions of social performance on firm risk. J. Bank. Finance **37**(4), 1258–1273 (2013)
Braun, A., Utz, S., Xu, J.: Are insurance balance sheets carbon-neutral? Harnessing asset pricing for climate change policy. Geneva Pap. Risk Insur. Issues Pract. **44**(4), 549–568 (2019)
Cao, J., Titman, S., Zhan, X., Zhang, W.: ESG preference and market efficiency: evidence from mispricing and institutional trading. SSRN Working Paper (2019)
Capelle-Blancard, G., Monjon, S.: Trends in the literature on socially responsible investment: looking for the keys under the lamppost. Bus. Eth. Eur. Rev. **21**(3), 239–250 (2012)
Capelle-Blancard, G., Monjon, S.: The performance of socially responsible funds: does the screening process matter? Eur. Financ. Manag. **20**(3), 494–520 (2014)
Carroll, A., Shabana, K.: The business case for corporate social responsibility. A review of concepts, research and practices. Int. J. Manag. Rev. **12**(1), 85–105 (2010)
Cheema-Fox, A., LaPerla, B.R., Serafeim, G., Turkington, D., Wang, H.: Decarbonizing factors. SSRN Working Paper (2019)
Davison, A., Kuonen, D.: An introduction to the bootstrap with applications in R. Stat. Comput. Stat. Gr. Newsl. **13**(1), 6–11 (2002)
de Carvalho Ferreira, M.C.R., Sobreiro, V.A., Kimura, H., de Moraes Barboza, F.L.: A systematic review of literature about finance and sustainability. J. Sustain. Finance Invest. **6**(2), 112–147 (2016)
DeMiguel, V., Garlappi, L., Nogales, F., Uppal, R.: A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. Manag. Sci. **55**(5), 798–812 (2009)
Derwall, J., Guenster, N., Bauer, R., Koedijk, K.: The eco-efficiency premium puzzle. Financ. Anal. J. **61**(2), 51–63 (2005)
Derwall, J., Koedijk, K., Ter Horst, J.: A tale of values-driven and profit-seeking social investors. J. Bank. Finance **35**, 2137–2147 (2011)
Dichtl, H., Drobetz, W., Wambach, M.: Where is the value added of rebalancing? A systematic comparison of alternative rebalancing strategies. Financ. Mark. Portf. Manag. **28**(3), 209–231 (2014)
Diltz, J.: Does social screening affect portfolio performance? J. Invest. **4**, 64–69 (1995)
Drempetic, S., Klein, C., Zwergel, B.: The influence of firm size on the ESG score: corporate sustainability ratings under review. J. Bus. Ethics (2019) **(forthcoming)**
Eccles, N.: UN principles for responsible investment signatories and the anti-apartheid SRI movement: a thought experiment. J. Bus. Eth. **95**, 415–424 (2010)
Eccles, N., Viviers, S.: The origins and meanings of names describing investment practices that integrate a consideration of ESG issues in the academic literature. J. Bus. Eth. **104**(3), 389–402 (2011)
Eurosif: European SRI Study 2018. Eurosif, Paris (2018)
Fama, E., French, K.: Size, value, and momentum in international stock returns. J. Financ. Econ. **105**(3), 457–482 (2012)
Fama, E., French, K.: A five-factor asset pricing model. J. Financ. Econ. **116**, 1–22 (2015)
Fama, E.F., French, K.R.: Choosing factors. J. Financ. Econ. **128**, 234–252 (2018)

Fox, J.: Applied Regression Analysis and Generalized Linear Models. Sage Publications, Los Angeles (2015)

Friede, G., Busch, T., Bassen, A.: ESG and financial performance: aggregated evidence from more than 2000 empirical studies. J. Sustain. Finance Invest. **5**(4), 210–233 (2015)

Friedman, M.: Capitalism and Freedom. University of Chicago Press, Chicago (1962)

Gabriel, K.: Ethik in der Geldanlage: Grundlagen, Kriterien und Herausforderungen. In: Scholz, S., Faust, M. (eds.) Nachhaltige Geldanlagen - Produkte. Frankfurt School Verlag, Strategien und Beratungskonzepte, Frankfurt (2014)

Galema, R., Plantinga, A., Scholtens, B.: The stocks at stake: return and risk in socially responsible investment. J. Bank. Finance **32**(12), 2646–2654 (2008)

Griffin, J.: Are the Fama and French factors global or country specific? Rev. Financ. Stud. **15**(3), 783–803 (2002)

Hahn, R.: Ethische Grundlagen des betrieblichen Nachhaltigkeitsmanagments. In: Baumast, A., Pape, J. (eds.) Betriebliches Nachhaltigkeitsmanagement. Verlag Eugen Ulmer, Stuttgart (2013)

Harjoto, M., Laksmana, I.: The impact of corporate social responsibility on risk taking and firm value. J. Bus. Eth. **151**(2), 353–373 (2018)

Harvey, C.R., Liu, Y., Zhu, H.:. and the cross-section of expected returns. Rev. Financ. Stud. **29**(1), 5–68 (2016)

Heinkel, R., Kraus, A., Zechner, J.: The effect of green investment on corporate behavior. J. Financ. Quant. Anal. **36**(4), 431–449 (2001)

Hong, H., Kacperczyk, M.: The price of sin: the effects of social norms on markets. J. Financ. Econ. **93**(1), 15–36 (2009)

Hudson, R.: Ethical investing: ethical investors and managers. Bus. Eth. Q. **15**(4), 641–657 (2005)

Humphrey, J., Tan, D.: Does it really hurt to be responsible? J. Bus. Eth. **122**, 375–386 (2014)

Humphrey, J.E., Lee, D.D., Shen, Y.: Does it cost to be sustainable? J. Corp. Finance **18**(3), 626–639 (2012)

In, S.Y., Park, K.Y., Monk, A.: Is 'being green' rewarded in the market? An empirical investigation of decarbonization and stock returns. SSRN Working Paper (2019)

Kanning, H.: Nachhaltige Entwicklung - Die gesellschaftliche Herausforderung für das 21. Jahrhundert. In: Baumast, A., Pape, J. (eds.) Betriebliches Nachhaltigkeitsmanagement. Verlag Eugen Ulmer, Stuttgart (2013)

Kempf, A., Osthoff, P.: The effect of socially responsible investing on portfolio performance. Eur. Financ. Manag. **13**(5), 908–922 (2007)

Kim, Y., Li, H., Li, S.: Corporate social responsibility and stock price crash risk. J. Bank. Finance **43**, 1–13 (2014)

Kotchen, M., Moon, JJ.: Corporate social responsibility for irresponsibility. BE J. Econ. Anal. Policy 12(1):Article 55 (2012)

Kuhlman, T., Farrington, J.: What is sustainability? Sustainability **2**, 3436–3448 (2010)

Lahiri, S.: Resampling Methods for Dependent Data. Springer, New York (2003)

Lai, T., Xing, H.: Statistical Models and Methods for Financial Markets. Springer, New York (2008)

Laurence, E.: Socially responsible investment. In: Idowu, S.O., Capaldi, N., Zu, L., Das Gupta, A. (eds.) Encyclopedia of Corporate Social Responsibility, pp. 2258–2263. Springer, Berlin (2013)

Ledoit, O., Wolf, M.: Robust performance hypothesis testing with the Sharpe ratio. J. Empir. Finance **15**, 850–859 (2008)

Lee, D., Faff, R.: Corporate sustainability performance and idiosyncratic risk: a global perspective. Financ. Rev. **44**(2), 213–237 (2009)

Lee, D., Faff, R., Rekker, S.: Do high and low-ranked sustainability stocks perform differently? Int. J. Account. Inf. Manag. **21**(2), 116–132 (2013)

Levitt, S.D., List, J.A.: What do laboratory experiments measuring social preferences reveal about the real world? J. Econ. Perspect. **21**(2), 153–174 (2007)

Lioui, A.: The decarbonisation factor: a new academic fiction? EDHEC Working Paper (2019)

Markowitz, H.: Portfolio selection. J. Finance **7**(1), 77–91 (1952)

Mattingly, J.E., Berman, S.L.: Measurement of corporate social action: discovering taxonomy in the Kinder Lydenburg domini ratings data. Bus. Soc. **45**(1), 20–46 (2006)

McNeill, J., Engelke, P.: The Great Acceleration: An Environmental History of the Anthropocene Since 1945. The Belknap Press of Harvard University Press, Cambridge (2014)

Merton, R.: A simple model of capital market equilibrium with incomplete information. J. Finance **42**, 483–510 (1987)

Patton, A., Politis, D., White, H.: Correction to automatic block-length selection for the dependent bootstrap by D. Politis and H. White. Econ. Rev. **28**, 372–375 (2009)
Pizzutilo, F.: Measuring the under-diversification of socially responsible investments. Appl. Econ. Lett. **24**(14), 1005–1018 (2017)
Politis, D., Romano, J.: The stationary bootstrap. J. Am. Stat. Assoc. **89**, 1303–1313 (1994)
Politis, D., White, H.: Automatic block-length selection of the dependent bootstrap. Econ. Rev. **23**, 53–70 (2004)
Pufé, I.: Nachhaltigkeit, 2nd edn. UVK Lucius, München (2014)
Renneboog, L., Horst, T., Zhang, C.: Socially responsible investments: institutional aspects, performance, and investor behavior. J. Bank. Finance **32**, 1723–1742 (2008)
Revelli, C., Viviani, J.L.: Financial performance of socially responsible investing: what have we learned? A meta-analysis. Bus. Eth. Eur. Rev. **24**(2), 158–185 (2015)
Sauer, D.: The impact of social responsibility screens on investment performance: evidence from the Domini Social Index and Domini Equity Mutual Fund. Rev. Financ. Econ. **6**(2), 137–149 (1997)
Scholtens, B.: Finance as a driver of corporate social responsibility. J. Bus. Eth. **68**(1), 19–33 (2006)
Schröder, M.: The performance of socially responsible investments: investment funds and indices. Fin. Mark. Portf. Manag. **18**(2), 122–142 (2004)
Schwegler, R.: Inrate's ESG impact rating methodology. https://www.inrate.com/cm_data/inrate_methodology_paper_newdesign_02.pdf (2018)
Statman, M., Glushkov, D.: The wages of social responsibility. Financ. Anal. J. **65**(4), 33–46 (2009)
Staub-Bisang, M.: Sustainable Investing for Institutional Investors: Risks, Regulations and Strategies. Wiley and Sons, Hoboken (2012)
Steffen, W., Sanderson, A., Tyson, P., Jäger, J., Matson, P., Moore, B., Oldfield, F., Richardson, K., Schellnhuber, H.J., Turner, B., Wasson, R.: Global Change and the Earth System. Springer, Heidelberg (2004)
United Nations Environment Programme Finance Initiative: Translating environmental, social and governance factors into sustainable business value. UNEP FI and World Business Council for Sustainable Development, Geneva (2010)
United Nations Principles for Responsible Investment: PRI Reporting Framework Main Definitions 2018. PRI Association, London (2017)
Utz, S.: Over-investment or risk mitigation? Corporate social responsibility in Asia-Pacific, Europe, Japan, and the United States. Rev. Financ. Econ. **36**(2), 167–193 (2018)
Utz, S., Wimmer, M.: Are they any good at all? A financial and ethical analysis of socially responsible mutual funds. J. Asset Manag. **15**(1), 72–82 (2014)
Van Marrewijk, M.: CSR and corporate sustainability: between agency and communion. J. Bus. Eth. **44**(2–3), 95–105 (2003)
Verheyden, T., Eccles, R., Feiner, A.: ESG for all? The impact of ESG screening on return, risk and diversification. J. Appl. Corp. Finance **28**(2), 47–55 (2016)
von Arx, U., Ziegler, A.: The effect of corporate social responsibility on stock performance: new evidence for the USA and Europe. Quant. Finance **14**(6), 977–991 (2014). https://doi.org/10.1080/14697688.2013.815796
Walker, T.J., Lopatta, K., Kaspereit, T.: Corporate sustainability in asset pricing models and mutual funds performance measurement. Fin. Mark. Portf. Manag. **28**, 363–407 (2014)
World Commission on Environment and Development. Our Common Future. United Nations (1987)
Zerbib, O.D.: A sustainable capital asset pricing model (S-CAPM): evidence from green investing and sin stock. SSRN Working Paper (2020)
Ziegler, A., Schröder, M., Rennings, K.: The effect of environmental and social performance on the stock performance of European corporations. Environ. Resour. Econ. **37**, 661–680 (2007)

Arnaud Gougler is a Ph.D. candidate at the Chair of Industrial Economics of the University of Fribourg and holds a M.A. HSG in Banking and Finance from the University of St. Gallen (HSG). He has several years of experience in the investment industry, working in the quantitative asset management field.

Sebastian Utz is an Assistant Professor at the School of Finance of the University of St. Gallen (HSG). His focus research area is sustainable finance. He has been published in journals such as Operations Research, European Journal of Operational Research, Journal of Risk, and Review of Financial Economics. Moreover, his research has been honored with the LIFE Climate Foundation Liechtenstein Impact Award and the Shin Research Excellence Award of the Geneva Association, among others.